Road network selection for small-scale maps using an improved centrality approach

Weiss, Roy E ; Weibel, Robert

Road Network Selection for Small-Scale Maps Using an Improved Centrality Approach

R. Weiss[1], R. Weibel[1]

[1]University of Zurich, Zurich, CH
Email: {roy.weiss | robert.weibel}@geo.uzh.ch

1. Introduction

The Federal Office of Topography of Switzerland (swisstopo) has the vision to ultimately derive their complete series of map products from a single, high resolution topographic landscape model, TLM3D. With a nominal scale between 1:5,000 and 1:25,000 (dependent on the area and feature class) TLM3D also serves as a basis to derive small-scale maps of 1:200,000 and smaller. Owing to the vast scale difference, the generalization process is currently still carried out completely manually in an interactive system. While the cartographers follow a specific set of rules, the process is heavily dependent on the experience of the particular cartographer.

In order to render the generalization process more repeatable and efficient, swisstopo is interested in automating the selection of certain feature classes, in particular the road network. They have thus inspired a research project which has the objective to develop a methodology for road network selection from a large-scale and detailed dataset (TLM3D) to a target scale of 1:200,000 and smaller.

According to swisstopo, several requirements need to be met by a road network selection algorithm, if it were to be used for the automatic derivation of their small-scale map. On such a map, the *importance* of a road in the network becomes the most determining criterion when thinning out the road network. As can be seen in Figure 1, only the most important roads should be kept, while preserving the general characteristics of the network (e.g. denser town centers should remain visible on the final map). In addition, there must not be disconnected parts and, as is evident in Figure 1, no new dead-ends should be created in the smaller scale map. While certain dead-ends exist, they often exist for a specific purpose only (e.g. to hint at an existing access road).

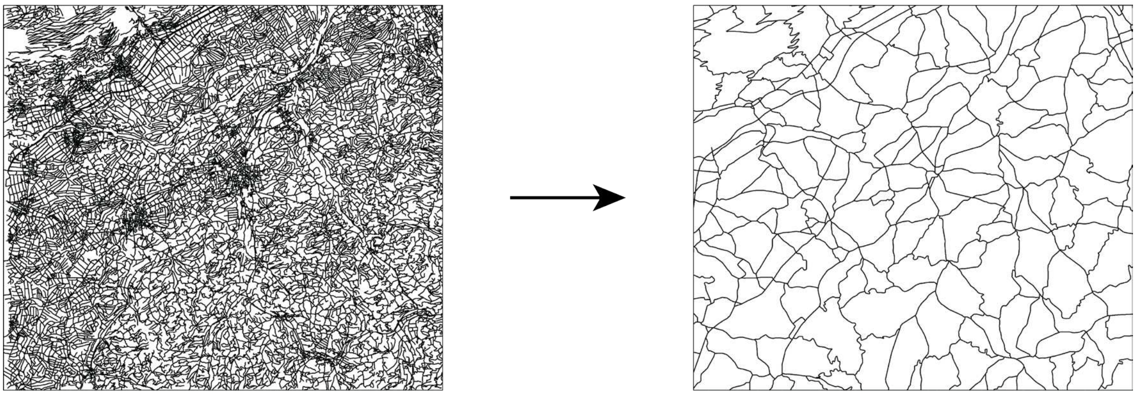

Figure 1. Test area of 26 km x 23 km around Langenthal in Switzerland of the source dataset (TLM3D) and the target dataset (VECTOR200). Data © swisstopo.

A review of the literature reveals that several interesting algorithms have been proposed for the problem of road network selection (Thomson and Richardson 1999, Jiang and Claramunt 2004, Touya 2010, Yang et al. 2011). Most notably, the stroke-based, graph-theoretic centrality approach by Jiang and Claramunt (2004) seemed very promising for road network selection in small-scale maps, as it incorporates the importance of roads. However, while the general principle and the results from earlier work looked like a perfect fit, it soon became apparent that several difficulties arise when dealing with larger areas (which becomes inevitable with a target scale of 1:200,000). Despite the fact that this approach has been proposed quite a while ago, it appears that it has not been tested on large areas of small-scale maps. This, however, is an indispensable criterion if the algorithm should ultimately be used in a production environment. Hence, a methodology on the basis of the approach proposed in Jiang and Claramunt (2004) was developed, which improved on the specific problems detected during the empirical phase of this work. This paper reports on the main elements of this methodology and the algorithms involved.

2. The Stroke-based Centrality Approach

The approach which serves as the main basis for the methodology was first proposed by Jiang and Claramunt (2004) and is based on graph-theoretic principles discussed by Mackaness and Beard (1993). In graph-theoretic approaches, the road network is seen as a graph, where the road segments represent the edges, while junctions (or endpoints of segments) represent nodes. Although it would be possible to build an algorithm on the segment level (i.e. on the primal graph of the road network), this does not represent the network really well compared to how humans see a road network. A cartographer does not assess the importance of individual segments, but of interconnected roads as a whole – the algorithm should do the same. It should not keep or eliminate individual segments, but whole roads. However, road names are not always available in databases and it is not always advisable to generate roads based on road names alone (Zhou and Li 2012). That is the reason why the approach builds upon so called strokes, which were introduced by Thomson and Richardson (1999).

Strokes group together individual road segments by means of the conceptual grouping principle of good continuation (Thomson and Brooks 2000). If translated to an algorithm, this essentially means that at each intersection, the angle between all segments is calculated and pairs of segments are grouped together based on their angle. In Figure 2, an example of such a concatenation is shown. For deciding which segments to concatenate, the algorithm examines the angle between each possible pair of segments and chooses the pair-combinations having the smallest angle in between. While the original proposal consisted of solely evaluating the angle of two segments, others have included the name of the road or other additional attributes, like the class of the road (Jiang and Claramunt 2004, Zhou and Li 2012).

After generating the strokes, a dual graph is created. In a dual representation of a graph, edges become nodes and vice versa (Porta et al. 2006). This means that each node represents a stroke, potentially consisting of multiple segments. The edges represent the intersections between the strokes. Such an example is shown in Figure 3, where the direct graph representation of a road network is shown on the left and its dual graph representation is shown on the right.

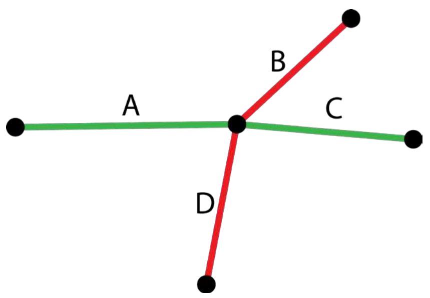

Figure 2. The best match for segment A is segment C, as C forms the smallest angle to A and vice versa. The same is true for B and D (based on Zhou and Li 2012).

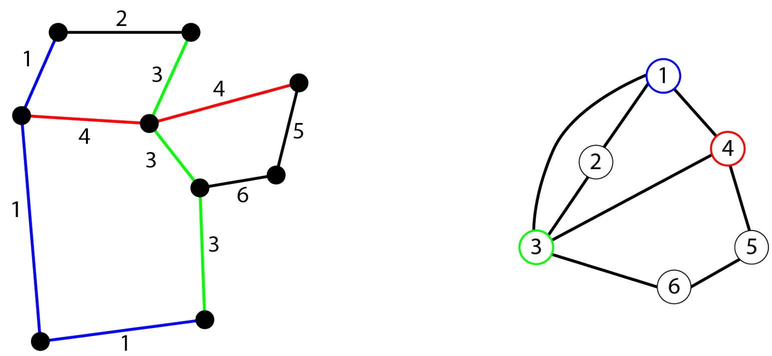

Figure 3. (A) The direct representation with colored strokes. (B) The dual representation with strokes as nodes (based on Porta et al. 2006).

Based on this dual graph, different centrality measures can be computed which describe and value a node's status in the network. Because nodes represent the generated strokes, they directly determine their importance. While there exist various centrality measures which were mainly developed for the analysis of social networks (Freeman and Linton 1978, Brandes 2001), the most useful one for the analysis of road networks is the so called *betweenness centrality*. This measure describes how often a node is used in a shortest path between two other nodes (Jiang and Claramunt 2004). Other measures, such as *closeness centrality* (the average distance of a node to all other nodes) and *degree centrality* (the direct links of a node) can also be used to further describe the rank of a road in a network. However, both closeness and degree centrality are not very reliable in determining the hierarchical importance of a stroke in a network (Yang et al. 2011). Betweenness centrality, on the other hand, can be used to identify roads which have a bridging role between different topological shortest paths, and thus in the entire network (Jiang 2009).

Various work has been done in the past using centrality measures to evaluate the importance of roads and ultimately thin out the network based on the calculated values. Jiang and Claramunt (2004) have introduced this approach by using a named-road approach to generate the strokes instead of concatenating the segments based on their deflection angle. Yang et al. (2011) extended the approach by identifying complex junctions and dual carriageways. They used the classical stroke approach using the deflection angle as the only criterion for the stroke generation process.

3. Problems and Proposed Solutions

Following the implementation of the original stroke-based centrality approach by Jiang and Claramunt (2004), various problems were identified in an initial series of tests. Most research so far has concentrated on relatively small, homogeneous areas, such as city centers. However, the problems that were detected become more serious for smaller target scales and larger, heterogeneous study areas. Clearly, the basic algorithm could not fulfill the requirements of swisstopo and small-scale maps in general. This section presents the problems that have been identified, as well as proposed enhancements to the algorithm to solve these problems:

- Usage of an improved stroke-generator that incorporates an enhanced set of semantic rules
- Detection and collapse of roundabouts
- Adaptive selection based on road density
- Intelligent reconnection of dead-ends and disconnected parts

3.1 Enhanced Semantic Rules for Strokes

There exist various strategies for the concatenation of segments into strokes. Jiang and Claramunt (2004) used road names, while Yang et al. (2011) solely used the deflection angle. Zhou and Li (2012) conducted a comparative study in which they evaluated the accuracy of these different approaches. They concluded that using road names is not advisable (incomplete acquisition, ambiguous results), but showed that using the deflection angle in addition to the road class provides the most accurate results.

As can be seen in Figure 4, the road class is in most cases more important than the deflection angle itself. In reality, the 4 m wide road makes a sharp turn which translates into the pictured situation. The concatenation in Figure 4a) makes less sense than the concatenation in Figure 4b) – thus, the road class comes first, the deflection angle second. In addition, the angle threshold should be higher than 60 degrees (an often used threshold) as there are many situations where a concatenation makes sense despite the high deflection angle. However, this does not mean that only segments of the exact same road class should be concatenated. Rather, the contrary is the case, as often a wide road flows into a narrower road or a narrow road gets temporarily wider. This means that some change in road class should be allowed. Therefore, the set of rules to apply in the stroke generation process becomes relatively complicated, as there exist many different road classes (14 in our case). Some of the classes should be allowed to connect to each other, while others should not. In addition, the angle threshold should be adjusted accordingly.

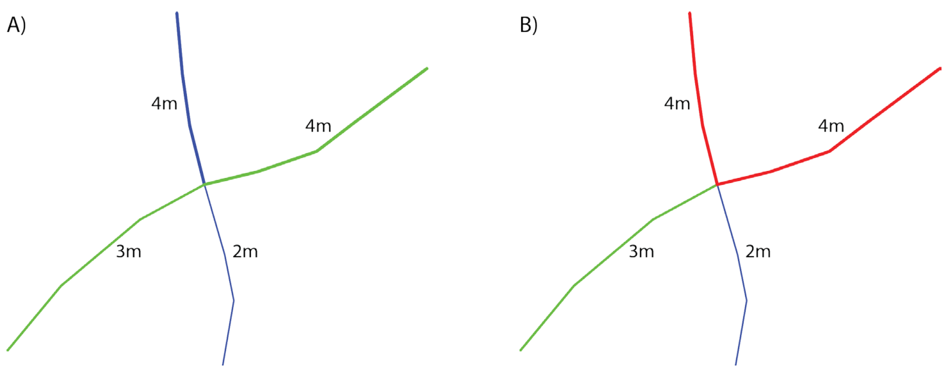

Figure 4. (A) Stroke concatenation using a simple deflection angle criterion. (B) Stroke concatenation using enhanced semantic rules.

In this work, the conclusion was that an angle of around 90 degrees is necessary to accommodate for cases like the one shown in Figure 4. However, several different angle thresholds were used for different cases. Especially smaller paths in mountainous areas benefited from the threshold increase, as very sharp turns happen quite often. This is sometimes also the case with wide roads: imagine a road leading up a mountain in hairpin bends. In such bends, it often happens that the road splits. Using 60 degrees as the threshold angle, the stroke would end at such positions, which is not a desirable result.

3.2 Detection and Collapse of Roundabouts

The main algorithm, as mentioned in Section 2, selects strokes based on some centrality measures. An accurate stroke building process is therefore essential. However, a main problem that arises are sections in the network where strokes are split up, like roundabouts or larger, very complex junctions. Yang et al. (2011) already proposed a solution using a clustering algorithm to detect such junctions with a subsequent correct concatenation of the strokes. Mackaness and Mackechnie (1999) also proposed a method to simplify junctions with clustering.

Based on our empirical research, we have found that the very large number of roundabouts in towns, rural regions and also cities break up many important strokes (see Figure 5). However, it is difficult to just detect roundabouts with clustering algorithms as they often either find too many clustered intersections or too few. In addition, the enhanced semantic rules already help to reduce the concatenation problems at complex junctions significantly. Hence, a different approach was developed to detect solely roundabouts, which does not rely on clustering.

Figure 5. Detected roundabouts in a small town in Switzerland. Data: TLM3D © swisstopo.

The reason for the splitting of strokes at roundabouts is obvious: as can be seen in Figure 6a, roads generally enter a roundabout in a very sharp angle. Based on the good continuation principle, the roundabout will not be connected to entering roads, but to itself – it gets isolated into a single, circular stroke when using the basic deflection angle rules without additional semantics. This drastically reduces the problem of identifying roundabouts in the network, as the strokes can be checked for loops individually.

Using some additional parameters which characterize a loop, the algorithm, developed together with Benz and Weibel (2013), detects roundabouts within strokes very effectively (all circular roundabouts in the test areas have been detected). After the detection and isolation of the roundabouts, their centroids are calculated and the incoming strokes are connected to the newly created nodes in the center, removing the roundabouts in the process (see Figure 6b).

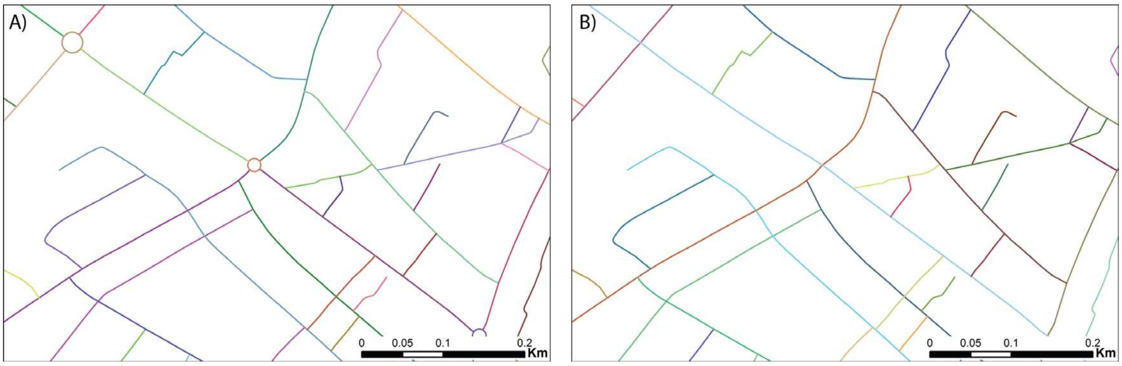

Figure 6. (A) Isolated roundabouts break up strokes. (B) Concatenated strokes after the detection and removal of roundabouts. Data: TLM3D © swisstopo.

3.3 Intermediate Result

Figure 7 shows the study area of Figure 1, after applying the centrality-based approach with the two improvements described in Sections 3.1 and 3.2 included. The comparison with Figure 1 shows that the main structures are well represented, though it should be pointed out that the manually produced VECTOR200 dataset is not restricted to selection operations, but also includes other generalization operators, such as road simplification and displacement. The selection result of Figure 7 also exhibits two problems, which are highlighted with ellipses: first, some parts of the network have become disconnected, creating dead-ends; second, in areas of high network density, the selected road network remained too dense (preventing the algorithm from using a higher threshold for less dense areas). Sections 3.3 and 3.4 present possible solutions to these problems.

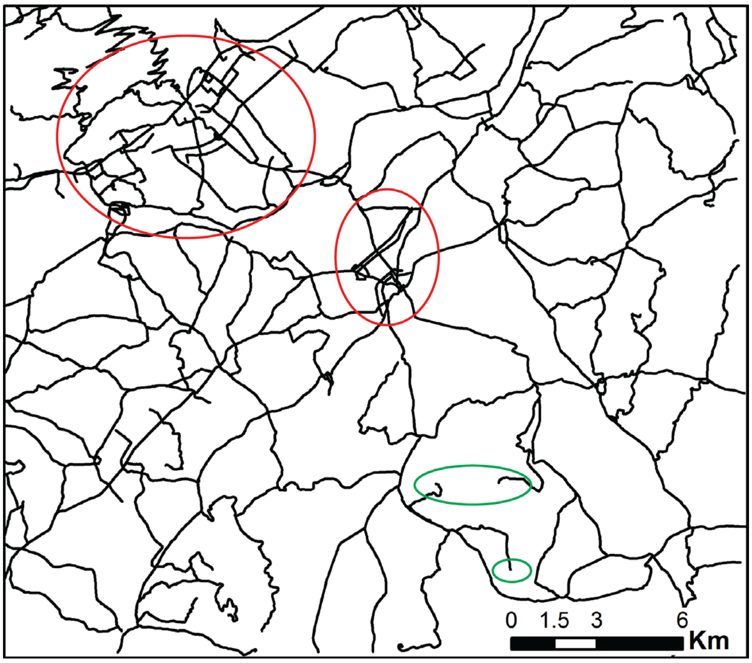

Figure 7. Pruned network of the study area with the improvements discussed in Sections 3.1 and 3.2. Two main problems still need to be tackled: certain roads have become dead-ends (green ellipses) and in areas of high density, the pruned network remains too dense (red ellipses). Data: TLM3D © swisstopo.

3.4 Adaptive Selection Based on Road Density

The centrality-based selection algorithm is able to identify important roads very well. However, betweenness centrality is tied to road density: when the road density in the original dataset is high, the strokes generally have a higher betweenness value. While this does not lead to problems in areas of homogeneous network density, as the threshold can simply be adapted, it leads to uneven pruning in heterogeneous areas. Towns and city centers, in relative terms, get thinned out less; rural parts of the area get more affected. While this is intended to some extent, the difference in density between rural and urban regions becomes too high. Hence, it is necessary to either use separate thresholds for dense areas (usually towns and cities) or remove specific roads in dense regions in a second pass.

The solution which has been implemented in this approach identifies dense regions after an initial selection has occurred. Afterwards, a DBSCAN (Ester et al. 1996) is performed and dense clusters are identified.

An example is illustrated in Figure 8. After the initial selection process has been completed, a dense region is extracted by clustering. The strokes in this specific area can then be handled separately in a next step. A different threshold is applied to the strokes contained in the dense region, which results in an adaptation of the road density in the dense regions. The implemented method offers an option to separately adjust thresholds for dense (usually urban) and less dense (usually rural) regions.

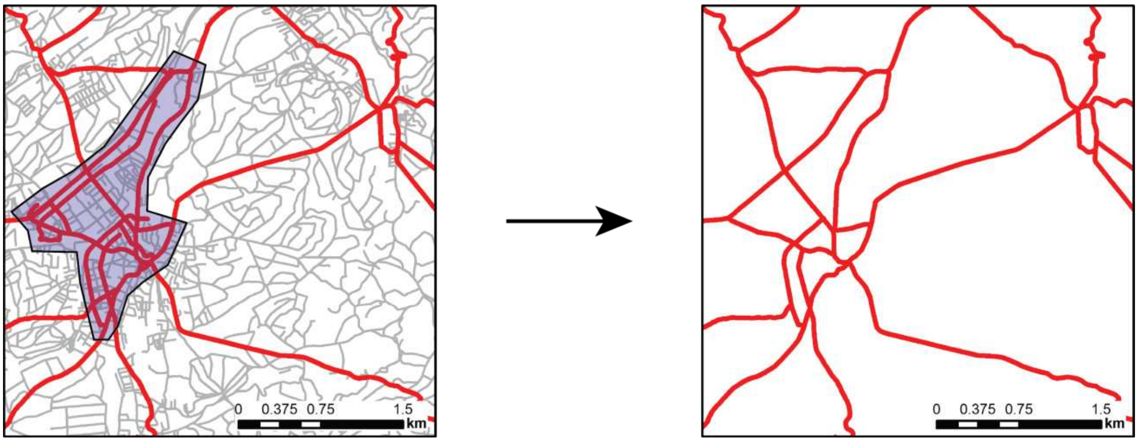

Figure 8. Detection of a high density area in the pre-selected network (red). A separate adjustment of the centrality thresholds in the specified high density area leads to an adapted selection. Data: TLM3D © swisstopo.

3.5 Reconnecting Dead-ends and Disconnected Parts

An important criterion for a usable, generalized map is that there are no disconnected parts in the road network. While this can be solved by reconnecting these parts using some kind of shortest path algorithm or by avoiding them in the first place using a Minimum Spanning Tree (Yang et al. 2011), it often happens that new dead-ends are created. However, as can be seen in Figure 1, where an extract of the actual road network of the 1:200,000 map of swisstopo is shown, nearly no dead-ends exist at that scale. Such a map should guide a user in finding directions – only under special circumstances should it split up roads or create new dead-ends, which in reality do not exist, for example, to show an access road to a facility.

A solution to this problem would be to either remove these parts completely or to reconnect them to the network. Obviously, this should not be done in a random way. The affected segments should get reconnected such that the network structure is taken into account, which is important for smaller scales. The connection between the main network and the road to be reconnected should therefore:

- be of a similar road class;
- lead in the same general direction; and
- be short.

The algorithm used generates all possible paths from the dead-ends back to the main network and choose the one with the best fit based on the above criteria. This way, both dead-ends and disconnected parts are reconnected to the network in a reasonable way and the utility of the pruned network is improved. Such an example is shown in Figure 9: the recursive search starts at the end node of the dead-end and creates all possible paths back to the network in the appropriate direction. The one with the best fit is then chosen and the connection to the main network is added to the pruned network.

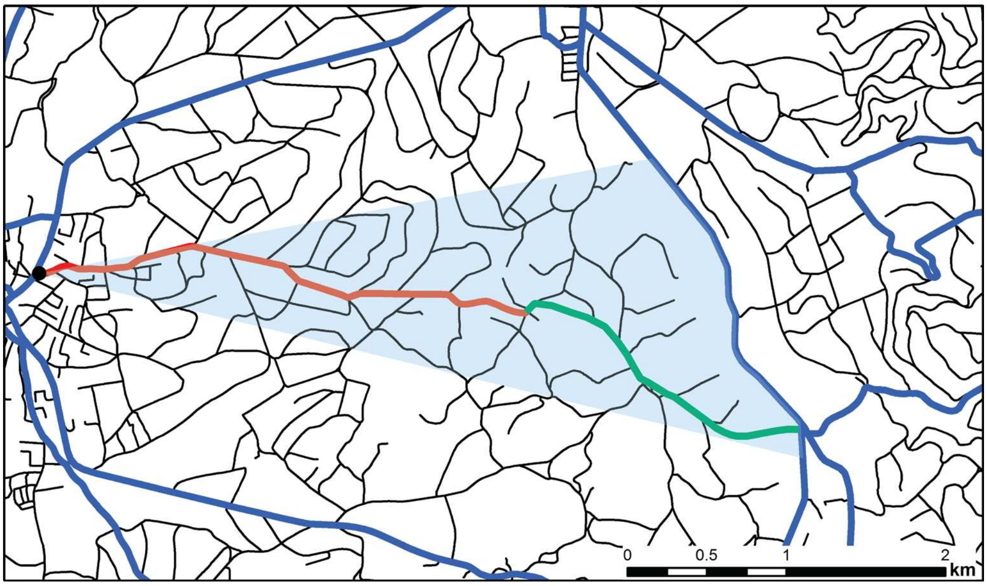

Figure 9. Initial pruning of the original network (black) generated a new network (blue) that contains a dead-end (red). To remove the dead-end, a recursive search through the original network is performed in the general direction (blue area) of the part to be reconnected, and the best solution (green), based on centrality, length and road class, is added to the pruned network. Data: TLM3D © swisstopo.

Figure 10 shows a possible result of one of the test areas. It is massively enhanced compared to the results obtained using a basic centrality approach. As can be seen, the dead-ends have either been removed or reconnected and the density in the urban regions reduced.

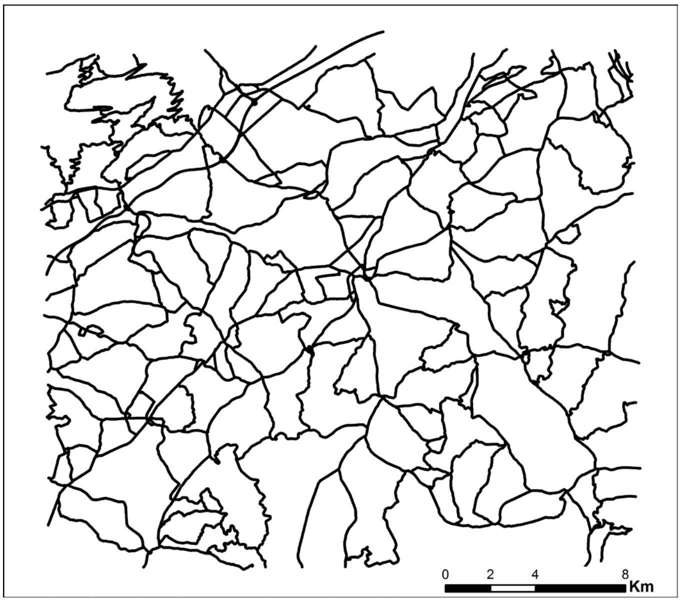

Figure 10. Possible final result of one of the test areas.

4. Conclusion and Future Work

The stroke-based centrality approach (Jiang and Claramunt 2004, Yang et al. 2011) has already shown good results in the past when generalizing road networks for medium-scale maps. For small-scale maps, however, there exist several problems, which hinder the approach from generating cartographically adequate road networks. In this paper, we have identified such problems and provided effective solutions to solve them. The approach has been tested on four different datasets of different characteristics. The prototype has been implemented in Java. The running time of the whole approach is reasonably fast: for a dataset of over 60’000 segments, the algorithm finished in less than 2 minutes on an Intel Q9550 CPU running at 3.4GHz with 8GB of memory available. A qualitative evaluation conducted by active cartographers of swisstopo has established that the adaptation of the basic algorithm to the needs of small-scale maps increases its usefulness significantly and the results resemble manually generalized maps more closely.

Nevertheless, the choice of appropriate thresholds for the centrality measures still requires further experimentation. While an appropriate threshold for a specific region can be found relatively quickly by empirical testing, it would be better if those thresholds were automatically generated, thus enabling the road selection methodology to work completely autonomously. More work still needs to be done to develop an effective algorithm for this particular problem.

Acknowledgements

We would like to thank swisstopo, in particular Cédric Métreaux and Dominik Käuferle, for providing the data as well as helpful inputs and feedback. Thanks are also due to Stefan A. Benz, who works on a related topic, for the long and fruitful discussions regarding the programming part of this work.

References

Benz SA and Weibel R, 2013, Road Network Selection Using an Extended Stroke-Mesh Combination Algorithm. In: *16th ICA Workshop on Generalisation and Map Production*, Dresden, Germany.

Brandes U, 2001, A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology,* 25 (2):163–177.

Ester M, Kriegel H, Sander J, Xu X, 1996, A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, USA, 226–231.

Freeman LC, 1978, Centrality in social networks conceptual clarification. In: *Social Networks,* 1(3):215–239.

Jiang B, Claramunt C, 2004, A Structural Approach to the Model Generalization of an Urban Street Network. *GeoInformatica,* 8(2):157–171.

Jiang B, 2009, Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science,* 23(8):1033–1048.

Mackaness WA, Beard KM, 1993, Use of Graph Theory to Support Map Generalization. *Cartography and Geographic Information Science*, 20(4):210–221.

Porta S, Crucitti P, Latora V, 2006, The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2):853–866.

Thomson RC, Richardson DE, 1999, The ‘good continuation’ principle of perceptual organization applied to the generalization of road networks. In: *Proceedings of the 19th International Cartographic Conference*, Ottawa, CA.

Thomson RC, Brooks R, 2000, Efficient Generalisation and Abstraction of Network Data Using Perceptual Grouping. In: *Proceedings of the 5th International Conference on GeoComputation*, University of Greenwich, London, UK.

Touya G, 2010, A Road Network Selection Process Based on Data Enrichment and Structure Detection. *Transactions in GIS*, 14(5):595–614.

Yang B, Luan X, Li Q, 2011, Generating hierarchical strokes from urban street networks based on spatial pattern recognition. *International Journal of Geographical Information Science,* 25(12):2025–2050.

Zhou Q, Li Z, 2012, A comparative study of various strategies to concatenate road segments into strokes for map generalization. *International Journal of Geographical Information Science*, 26(4):691–715.